Table 1: Experimental results on 6 real data sets from UCI Machine Learning Repository. For each pair ⟨task, algorithm⟩ an average testing error obtained from 10-fold cross validation is given, in percents. For each task four best results are bold-emphasized. Algorithms 1–7 are baseline rule learners. Our algorithms: WV — Weighted Voting, DL — Decision List, CB — Committee Boosting, SC — using heuristic modified by SC-bound, MC — using heuristic modified by Monte-Carlo estimation of overfitting.

	algorithms	tasks					
		australian	echo-card	heart dis.	hepatitis	labor	liver
1	RIPPER−opt	15.5	**2.9**	19.7	20.7	18.0	32.7
2	RIPPER+opt	15.2	5.5	20.1	23.2	18.0	**31.3**
3	C4.5 (Tree)	14.2	5.5	20.8	18.8	14.7	37.7
4	C4.5 (Rules)	15.5	6.8	20.0	18.8	14.7	37.5
5	C5.0	**14.0**	4.3	21.8	20.1	18.4	31.9
6	SLIPPER	15.7	4.3	19.4	**17.4**	**12.3**	32.2
7	LR	14.8	4.3	19.9	18.8	14.2	32.0
8	WV	14.9	4.3	20.1	19.0	14.0	32.3
9	DL	15.1	4.5	20.5	19.5	14.7	35.8
10	CB	**13.8**	**2.4**	**19.3**	21.4	**10.9**	32.3
11	**WV+MC**	**13.9**	3.0	19.5	**18.3**	13.2	**30.7**
12	**DL+MC**	14.5	3.5	19.8	18.7	13.8	32.8
13	**CB+MC**	**14.0**	**2.3**	**18.9**	19.9	**8.9**	**31.4**
14	**WV+SC**	14.1	3.2	**19.3**	**18.1**	13.4	**30.2**
15	**DL+SC**	14.4	3.6	19.5	18.6	13.6	32.3
16	**CB+SC**	14.9	**0.9**	**18.5**	**18.0**	**11.9**	42.7

Experiment. We use state-of-the art algorithms C4.5 (**?**), C5.0 (**?**), RIPPER (**?**), and SLIPPER (**?**) as baseline rule learners. Our rule learning engine is based on breadth-first search as features selection strategy. Fisher's exact test (**?**) is used as heuristic H. To build compositions of rules we use three algorithms. Logistic Regression (LR) is a linear classifier that aggregates rules learned independently. Weighted Voting (WV) is a boosting-like ensemble of rules, similar to SLIPPER, which trains each next rule on reweighted training set. Decision List (DL) is a greedy algorithm, which trains each next rule on training objects not covered by all previous rules. Committee Boosting (CB) is again a boosting-like ensemble of rules where all rules contribute with the same weight. In learning phase Committee Boosting differs from WV by means of ensuring rules diversity: to learn each rule it selects train subsample based on object's margins instead of reweighting objects.

There are two modifications of heuristic $H'(p, n)$. The SC-modification uses SC-bound on the probability of overfitting Q_ε as described above. The MC-modification uses the Monte-Carlo estimation of Q_ε via 100 random partitions $\mathbb{X} = X \sqcup \bar{X}$. For both modifications we set $\ell = k$.

In all experiments with CB we used a different sampling technique than we used for DL and WV. While in DL and WV setting we used theorem **??** and sample low layers of SC-graph, for CB we only used classifiers generated by our greedy base rule learner.

Table 1 shows that initially our algorithms WV, DL and CB are comparable to the baseline. WV outperforms DL, which corresponds to the results of other authors. Both SC- and MC- modifications reduce overfitting significantly and always outperform their respective initial versions. The difference between SC- and MC- modifications is not significant. Then, a moderate looseness of the SC-bound does not reduce its practical usefulness as a rule selection criterion.